

WHAT IS CLAIMED IS:

1. A method for relating words in an audio file to words in a text file, comprising:

retrieving a text file comprising a plurality of textual words;

generating an audio file comprising a plurality of audible words based on the text file; and

storing information relating each audible word to a corresponding textual word.

2. The method of Claim 1, wherein the textual words comprise ASCII text.

3. The method of Claim 1, wherein the audio file is stored in the form of a WAV file.

4. The method of Claim 1, wherein the information comprises voice tags embedded in the audio file.

5. The method of Claim 1, wherein the information comprises a file map relating a location of each textual word within the text file to a location of the corresponding audible word in the audio file.

6. The method of Claim 1, wherein the steps of the method are performed by logic embodied in a computer readable medium.

7. A method for relating words in an audio file to words in a text file, comprising:

retrieving a text file comprising a textual word;

generating an audible word corresponding to the textual word;

storing the audible word in an audio file;

storing a file map, the file map comprising:

a first location locating the audible word within the audio file; and

a second location locating the textual word within the text file.

8. The method of Claim 7, further comprising repeating the steps of the method for a plurality of textual words in the text file.

9. The method of Claim 7, further comprising:
receiving a command from a user to spell the audible word;

determining that the textual word corresponds to the audible word; and

audibly spelling the textual word.

10. A method for relating words in an audio file to words in a text file, comprising:

retrieving a text file comprising a plurality of textual words;

generating an audible word corresponding to each textual word, each audible word comprising media stream packets; and

playing the audible words to a user in real time as the audible words are generated; and

during the playing of the audible words, determining a current textual word corresponding to the audible word currently being played.

11. The method of Claim 10, wherein the textual words comprise ASCII text.

12. The method of Claim 10, further comprising:
initializing a counter identifying textual words within the text file; and

incrementing the counter after each audible word is played;

wherein the step of determining comprises identifying the current textual word using the counter.

13. The method of Claim 10, further comprising:
after each audible word is played, storing information about the audible word, the information comprising:

an identifier for the textual word corresponding to the audible word; and

a time at which the audible word was played.

14. The method of Claim 10, wherein the steps of the method are performed by logic embodied in a computer readable medium.

15. A method for relating words in an audio file to words in a text file, comprising:

retrieving a text file comprising a textual word;

generating an audible word based on the textual word,
5 the audible word comprising media stream packets; and

storing an identifier for the textual word.

16. The method of Claim 15, further comprising
repeating the steps of the method for a plurality of textual
words in the text file.

17. The method of Claim 15, further comprising:
receiving a command from a user to spell the audible
word;

determining that the textual word corresponds to the
audible word; and

audibly spelling the textual word.

18. A method for audibly spelling a word in an audio file, comprising:

playing an audio file to a user;

receiving from the user a command to spell an audible
5 word in the audio file;

identifying in a text file a textual word corresponding to the audible word; and

audibly spelling the textual word.

10 19. The method of Claim 18, wherein receiving the command comprises receiving a barge-in command during the playing of the audio file, and the method further comprises:

stopping the playback of the audio file;

15 identifying the last word played before the barge-in command was received; and

selecting the last word played as the audible word to be spelled.

20 20. The method of Claim 19, further comprising:

receiving a command from the user to resume playing the audio file; and

playing the audio file from the point at which playback was stopped.

25 21. The method of Claim 18, further comprising:

receiving a command from the user to select a new textual word from the text file; and

audibly spelling the new textual word.

22. An interactive voice response server (IVR),
comprising:

an interface operable to play an audio file to a user
and further operable to receive a command to spell an audible
word in the audio file from the user;

a processor operable to:

identify an audible word to be spelled in response
to the command to spell;

identify a textual word in a text file
corresponding to the audible word to be spelled; and
audibly spell the textual word.

23. The IVR of Claim 22, further comprising an adaptive
speech recognition (ASR) module operable to:

receive speech from the user; and

parse the speech into recognizable grammar, words or
vocabulary.

24. The IVR of Claim 22, wherein:

the interface is further operable to receive a command
from the user to resume playing the audio file; and

the processor is further operable to resume playing the
audio file in response to the command.

25. The IVR of Claim 22, wherein:

the interface is further operable to receive a command
to select a new textual word from the text file; and

the processor is further operable to select and to
audibly spell the new textual word.

26. Logic embodied in a computer readable medium operable to perform the steps of:

playing an audio file to a user;
receiving from the user a command to spell an audible
word in the audio file;
identifying in a text file a textual word corresponding
to the audible word; and
audibly spelling the textual word.

27. The logic of Claim 26, wherein receiving the command comprises receiving a barge-in command during the playing of the audio file, and the logic is further operable to perform the steps of:

stopping the playback of the audio file;
identifying the last audible word played before the barge-in command was received; and
selecting the last audible word played as the audible word to be spelled.

28. The logic of Claim 26, wherein the logic is further operable to perform the steps of:

receiving a command from the user to resume playing the audio file; and
playing the audio file approximately from a point at which playback was stopped.

29. The logic of Claim 26, wherein the logic is further operable to perform the steps of:

receiving a command from the user to select a new textual word from the text file; and
audibly spelling the new textual word.

30. A text-to-speech (TTS) system, comprising:

a memory operable to store a text file and an audio file; and

a TTS module operable to:

5 generate an audible word corresponding to each textual word in the text file;

store the audible words in an audio file; and

store for each audible word:

a first location locating the audible word in the audio file; and

a second location locating the corresponding textual word in the text file.

31. The system of Claim 30, wherein the system further comprises:

an output device operable to play the audio file to a user;

an interface operable to receive a command to spell one of the audible words during the playing of the audio file; and

a processor operable to:

determine the textual word corresponding to the audible word to be spelled; and

audibly spell the textual word.

32. Logic embodied in a computer readable medium, comprising:

selecting a textual word in a text file;

generating an audible word corresponding to the textual word;

storing the audible word in an audio file;

storing a file map, the file map comprising:

a first location locating the audible word within the audio file; and

a second location locating the textual word within the text file.

33. The logic of Claim 32, further operable to repeat the steps for a plurality of textual words in the text file.

34. The logic of Claim 32, further operable to perform the steps of:

receiving a command from a user to spell the audible word;

determining that the textual word corresponds to the audible word; and

audibly spelling the textual word.

35. A method for synchronizing audible words with textual words in a text file, comprising:

retrieving a text file comprising a plurality of textual words;

5 generating a plurality of audio files, each audio file comprising an audible word corresponding to one of the textual words; and

for each audio file, storing information relating the audio file to the corresponding textual word.

10 36. The method of Claim 35, wherein the steps are performed by logic embodied in a computer readable medium.

37. A system for spelling words in an audio file,
comprising:

means for playing an audio file to a user;

5 means for receiving from the user a command to spell an
audible word in the audio file;

means for identifying in a text file a textual word
corresponding to the audible word; and

means for audibly spelling the textual word.

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32